

GAPS, NEEDS, AND OPPORTUNITIES

As the findings of the literature illustrate, relevant articles exist across several domains; this means, among other things, that the research reflects different values, styles of argument, and evidentiary standards and conventions. Further, as the discussion of intellectual property in the Findings section illustrated, some of the concerns in the literature represent special cases or subsets of a broader range of concerns—which can be, in some instances, highly charged. Decisions affecting scientific journal publishing may be made for any number of reasons external to this enterprise, yet may profoundly affect the conditions under which scientists function.⁵⁴ All of these factors have inhibited generalization but contribute to the richness of the literature; moreover, they present many challenges for future research.

We posited five questions to guide this study:

- What issues arise from the literature?
- How do information scientists measure “impact” or implications or effects?
- Have changes in researchers’ behavior been discerned?
- What are the implications for underserved populations in the United States or abroad?
- Are information security (that is, how systems and data are protected from unauthorized use) and user privacy investigated?

The material presented in the Findings section suggests that there are no easy answers to any of these questions. Nonetheless, summary remarks pointing to areas for future research are offered below.

⁵⁴An obvious example is the dispute over Napster, a system for sharing music files over the Web. Although the Napster site posted careful copyright protection notices, use of the Napster system was believed to contribute to widespread infringement of rights. The issues that have been raised by the case may speak to fundamental issues of copying files, the status of search services, and—more generally—actions that scholars consider within “fair use.” Similar concerns have been voiced with respect to software reengineering and the requirements of the Digital Millennium Copyright Act and the Uniform Computer Information Transactions Act.

ISSUES ARISING IN THE LITERATURE

The literature is dominated by discussions over the relationship between the formal, peer-reviewed e-journal article and the larger hierarchy of scholarly and scientific communication forms and the extent to which the new information technologies have altered and disrupted traditional roles. This affects issues as fundamental as how the notion of an “e-journal” or “e-publication” is even defined, complicating efforts to determine the numbers of e-journals and to understand their implications. Some observers have seen in this situation the possibility for reform of the publishing system: they cite proposals that include eliminating or reducing the role of publishers; changing or eliminating peer review, which has historically been a function coordinated by journal publishers but has a strong element of being a “public good”; and changing how intellectual property rights are managed. Essential to any future research is clarification of what is being studied (i.e., the electronic journal or electronic journal article), whether the entity exists in multiple formats, is subject to peer review and formal editing, and is destined for formal archiving, which affects perceptions of reliability and availability for future investigators.

Associated with this debate is the question of pricing electronic journals, which is part of a general discussion among economists about methods for pricing information goods and which has also become embedded in the concern among librarians over escalating prices for serials. This too represents an area of emerging research, and, as of this writing, early experimental results are just beginning to be released. Pricing models require assumptions about how e-journals will be used and valued—an area in which results are still preliminary, diffused, and evolving as e-journals come to be more widely accepted (albeit at varying rates among scientific fields). Pricing strategies may also vary depending on whether the journal is available in multiple formats and is included as part of a professional membership, for example.

Other relevant issues, possibly less contentious but equally interesting, concern the functions and attributes of the e-journal article, or the new artifact; the relationship between the electronic and print artifacts (there is a decided preference for retaining some form of print); and archiving. Given the heterogeneity of the literature, the clear cultural differences in communication practices

among the sciences, and the associated variation in diffusion of the information technologies within the sciences—which is affected by the nature of the research as well as by traditions of formal and informal communication—it is not surprising that there is no obvious consensus on what has transpired and still less on what is likely to transpire. Indeed, there is even debate as to whether the changes are evolutionary and incremental or revolutionary and transformative, which is part of a long tradition that pits technological determinism at one extreme (i.e., once a technology is introduced, its ability to transform social relations is a question of time) and a view that emphasizes social organization, which argues that social organizations are based on normative order and their ideological underpinnings to which change is presumably subordinate (see Walsh and Bayma 1996).

Existing surveys largely from the mid-1990s indicate that researchers are generally content with existing codes of peer review and copyright practice. These attitudes may, however, change as debates over remote and online higher education and ownership of course material intensify and take on concrete economic value. Thus, continued study of attitudes, employing different research methodologies, including but not limited to observation, surveys, and anecdotal reports, is important—particularly if the studies couch the investigation not solely in terms of e-journal publishing but more broadly to take into account the way that higher education and research are being transformed. Commercial influences, documented by Walsh and Bayma (1996, 1997), may have broader effects than those associated with formal publication, but increased awareness of potential economic advantage might conceivably provoke reevaluations of such apparently accepted practices as transfer of copyright. At the same time, as noted by Kling and his colleagues, the behavior of key publishers can have a significant impact. This was evident in the roles played by the American Mathematical Society and the American Physical Society in encouraging their members to adopt the new technologies. The continued influence of associations and societies in determining what is acceptable professional conduct is worth investigation.⁵⁵

The kinds of discussions that occur within the literature reflect realignment of roles and functions, which is

a condition that has characterized other contexts in which information technology has been introduced. This process of realignment is exacerbated by ongoing concerns for the journal process, which predates the information technologies, and by a period of technological experimentation. However, one of the implications of the literature that is not directly expressed is the effect of transparency, which results from the application of information technologies and from changes that these applications inspire. That is, costs that were once absorbed become obvious, and there arises an opportunity to model the relationships anew and in the context of a realignment of institutional roles and functions that the technology is abetting, if not provoking. This issue underlies the renewed interest in archiving and peer review and, more generally, in a “public goods” model of scientific publishing.

As noted above, Berry (2000) makes the public goods case; that is, scientific research is supported by funds from government and not-for-profit agencies and the value of its findings is not *diminished* by use but, because science is cumulative, is *increased* by use. Consequently, conventional market models do not apply. There is an additional “public good” argument which Berry does not fully make although others, including Hal Varian, have. The public good argument centers on the concept of “non-appropriability” of benefits. Specifically, the producer cannot “appropriate” an adequate share of the benefits to recoup his costs so he produces less than is societally optimal. Thus, the “market” fails to yield an optimal solution, and there may be a role for public intervention through the funding of dissemination, as Berry argues, or through a new program or third-party accrediting service. This is an interesting avenue for future research and potentially affects the ways peer review and archiving might be modeled and eventually priced within accepted frameworks for public goods and public interests.

One of the functions of print journal publishers is to coordinate peer review, and several proposals for reform call for variants of bottom-up or community-based evaluation of the value of individual pieces of work. There appears to be no consensus on the components of the publishing model, although various approaches have been set forth; of these, Tenopir and King’s, which relies heavily on print, appears to be the most complete. Other models (e.g., Peters 1998) deserve careful scrutiny. In general, there is surprising variability in the way that publishing and publishing costs are modeled.

⁵⁵In this regard, note that the American Physical Society is undertaking a study of publishing and that the journal *Nature* hosts an online forum on electronic publishing (<http://www.nature.com/nature/debates/e-access/index.html>).

A particular sticking point in this modeling is accounting for costs of coordination and of time volunteered by authors, reviewers, and editors, which is frequently set at essentially zero rather than estimated. Although these costs may be absorbed by the system or may be considered “matching” or displaced costs in some approaches, they are—regardless of whether they are directly embodied in the price of the object—nonetheless real (see Bergstrom 2001). Because there is a societal interest in peer review—that is, the validation, through explicit and accountable channels, of the research results on which public decisions are made whether in medicine or nuclear energy—modeling it from a public goods perspective would appear to be an interesting approach to this issue. This public goods approach is different from—although complementary to—a “ground up” approach (as proposed by Varian among others), since the peer review function would be institutionalized. Its institutionalization would make peer review accountable in a way that informal, self-organizing commentary that might be taken to represent the consensus of the community is not.

Archiving, which is related to the overall acceptance of e-journals, is another area that is currently being studied from a technological perspective but not from an economic perspective. Archiving is a rather complex issue, since electronic archives, unlike their print counterparts, require active management. Storage physically degrades, and, more importantly, the software systems that enable the data to be read become obsolete. An electronic archive must be continually “refreshed” if it is to remain useful. Who will archive, or pay for archiving, whether these costs can be tied into the access costs of current journals or should be assigned to the backfile (as is the case with both *The New York Times* and *The Washington Post*), is unclear. It is likely that much will depend on technology. Nevertheless, the organizational and economic questions associated with archiving require investigation—again, possibly from a public goods perspective.

Clearly, there is substantial work to be done in the area of pricing and its relationship to behavior as well as in modeling the costs of the e-journal system and its relationship to the larger spectrum of scientific communication. McCabe (2000) recommends consideration of the monopolistic market characteristic of scientific publishing from the perspective of antitrust: comparative examination of STM fields, the behavior of nonprofit publishers, and the entry of new journals into the market. More work on behavioral issues also appears promising, as evidenced by the PEAK project, since any attempt to

price e-journals (or any information good for that matter) presumably requires a better understanding of what the user wants or will accept. An interesting avenue for investigation might be to look at which of the many potential attributes of the new artifact have commercial value in the context of developing the model of differential pricing advanced by Varian. These approaches might also note what features are considered desirable, so that pricing strategies take the evolution of the artifact into account.

A final key issue in the future of e-journal publishing is intellectual property rights. The authors of the *Digital Dilemma* report (CSTB 2000) summarize four general areas for future research in intellectual property rights, all of which bear to some degree upon issues specific to STM e-publishing: the extent of illegal copying; the economics of copyright, patents, and “Cyber Law;” the validity of maintaining a legal regime with “copy” as the foundational concept; and the relationship between “content creators” and the digital environment. The extent to which each of these affects patterns and processes of scientific e-communication and publishing remains to be seen.

HOW INFORMATION SCIENTISTS MEASURE IMPACT

Impacts of individual articles or journals are typically measured through well-understood bibliometric techniques. Efforts are being made to extend this methodology to hyperlinks and to extend the notion of citations to include broader acknowledgment of other ways that impacts might be expressed. These efforts derive directly from the nature of the Internet/Web environment, which affords both multiple formats for providing information (home pages, white papers, electronic versions of peer-reviewed journal articles, listserv discussions, and so on), as well as the means for establishing explicit connections between and among information sources through hyperlinks. Assessing the impact of electronic publications as well as expanded methods for measuring impact are areas of study that are still in their infancy and which appear to be very promising.

Another as yet unexplored area for impact analysis is to measure user acceptance of information technologies, particularly in terms of the implications of such acceptance in understanding the structure of scientific communication. Measures of acceptance also, and obviously, reflect on what users want and therefore affect pricing.

The significance of other forms of electronic communication, most notably electronic preprints, also remains to be determined. As several observers have maintained, there are substantial differences among various science communities in the use and recognition accorded to preprints/e-prints; there likely may be variation across communities in the way their impact should be assessed. For example, the importance of e-print servers to the high energy physics community has been well documented, but many of the social sciences have yet to develop a similar mechanism. Cross-disciplinary studies based on communication styles and traditions within and among the various communities (for example, between mathematics and history or political science) are recommended.

CHANGES IN RESEARCHERS' BEHAVIOR

There is evidence of changing behaviors and attitudes toward electronic publications, as well as toward peer review, which are necessary in order for e-publications to gain credibility. Again, the studies are highly heterogeneous in design, scale, and rigor, making comparisons and the ability to draw conclusions difficult. Moreover, several of the larger scale studies have been conducted within the framework of the U.K. system of higher education; differences between expectations here compared with the United States should, at a minimum, be explored.

Larger scale studies with larger samples that span several disciplines and institutions and that employ a range of methodologies (e.g., quantitative, interviews, observation, ethnographic) are clearly needed. Such studies should examine differential penetration of the technology as well as use of electronic publications across various scientific fields. They might well elucidate variations among the various "subcultures" of scientific research communities along the lines developed by Walsh and Bayma as well as by Kling and his colleagues and should be consistent with the longitudinal study under way at the University of California–Los Angeles (see <http://ccp.ucla.edu/pages/InternetStudy.asp>). If possible, and based on the nature of the questions, definition of different levels of computer usage is recommended. While all researchers can read and the Web appears to be essentially ubiquitous within the research community, not all natural and social scientists read electronic journals or use computing-intensive analysis and methods as a basic

research tool. Thus, studies that seek to understand how electronic publication is different from other analog forms of publication and communication should consider differential use of computing as well as differential use of e-publications. Moreover, the studies to date clearly demonstrate differences among the ways that scientists use electronic journals for research, discovery, and browsing and in the venues in which they are comfortable publishing, particularly for purposes of promotion and tenure.

Few of the studies reviewed identified a control group of nonusers, and many of them are vulnerable to the self-selection bias that was occasionally acknowledged by the investigators. A number of studies complemented the survey questionnaire with follow-up interviews and focus groups in an effort to eliminate or at least mitigate this bias. Acceptable sample sizes vary from domain to domain, and it is obvious that the observation-intensive, diary-style studies used by computer science researchers, for example, are self-limiting, where the intensity and depth of the observations are balanced against the small numbers of participants. On the other hand, small studies of faculties are useful.

Comparisons among these different approaches are tenuous, in part because the designs are not necessarily comparable and in part because the results are so disparate. The collection of small-scale studies examined offers a useful source for developing appropriate designs for further, more broadly conceptualized projects. Such future projects might span domains that are interdisciplinary (e.g., oceanography), computing intensive (e.g., genetics) or noncomputing intensive (e.g., psychology), and employ multiple methodologies (e.g., ethnographic, quantitative, descriptive). Additionally, since early user studies indicated that one factor contributing to the adoption of e-journals was critical mass of information, future studies might do well to be cognizant of changes leading to expansion in the availability of e-journals as well as the expansion in numbers of potential users. Thus, time has a possible effect, and future investigators should take it into account in designing their projects.

Some other ideas that have been touched on in prior work but that might merit expanded inquiry include the following:

- *Reading behaviors:* Several studies have looked at the amount of time researchers spend reading, but this question might be usefully parsed into the kinds of reading that is done at different points in

people's careers (as a student, Ph.D. candidate, or research assistant; by various ranks of professor; during research in corporate labs; and so on) and across fields. To build on and supplement research done by Walsh, Kling, Odlyzko, and others, it might be productive to understand reading in the context of the research structure—i.e., whether the work is conducted in teams (e.g., as in biology, chemistry, computer science, and experimental physics) or on a more solitary basis (e.g., as in mathematics or economics). Other considerations that might be studied with regard to reading behaviors include the effect of the market, internationalization, the kinds of information sought, and the existence of related and supporting material in digital form.

- *The hierarchy or continuum of scientific communication:* Clearly, information technologies have afforded a wider range of communication modes, and although the journal article—whether electronic or print—remains a critical factor in promotion and tenure decisions, the extent to which other forms of communication (preprints, technical papers, conference papers, etc.) come to be recognized is interesting. While in the past it may have been difficult to track the influence of conference papers, as more of them are put on the Web and as the techniques for capturing influence relationships via the Web are evolved, it may be possible to measure other forms of influence and impact. In time, these may come to affect decisionmaking, particularly with respect to promotion and tenure. Again, studies that focus on variation within and across fields will be important. Odlyzko has pointed to the implications of preprint servers in physics and mathematics. It is unclear whether the same system can or will be replicated across all the sciences, particularly where there are different modes of research, work practices, and traditions of collaboration. The importance of variation between the (natural) sciences and social sciences, for example, has been well documented in the SuperJournal project.

Note that this kind of research, which relies in part on the expanded notions of “acknowledgment” or “invocation” as previously described, poses new considerations for personal privacy. The SuperJournal project has also grappled with the privacy implications of conducting research

online. Privacy is a core value in the research community, as is the creation of new knowledge. However, the capabilities of the new information technologies to support new kinds of communication and research can have the unintended consequence of pitting established values against each other.

- *Authoring behaviors:* There are limited-scope studies on the willingness of authors to write for new media. This type of study could be expanded by looking at a progression of author behaviors over the course of their careers and how these patterns differ by field. Affecting the decision to publish in electronic media are not only the culture of the particular science field but also issues of patentability, time to market, seniority, penetration of the information technologies into the conduct of the research (e.g., the importance of the genomic and protein sequence databases, visualization and scenarios for testing alternative hypotheses), and critical mass of information.
- *Change over time:* There is evidence that behavior with respect to technology changes as users become more familiar with it. Thus, longitudinal studies that capture users' comfort level and the interaction between users and their contexts (professional, institutional, etc.) might be extremely productive. What happens to the behavior of individual researchers, for example, as the disciplines become more heavily invested in the information technologies and analysis predicated on the capabilities of technology becomes more widely accepted? Genetics and molecular biology seem to be natural starting points, and the social sciences in general may be ripe for analysis of these kinds of questions; this might be particularly true for economics, where computer-assisted simulation is a promising tool.
- *Institutional relationships:* The policies of the major journals as well as of the leading scientific professional associations have substantial effects upon scientists' behavior. For example, the American Physical Society and American Mathematical Society played important roles in the development of standards, as has the National Science Foundation. Capturing the implications of these kinds of “environmental” factors is challenging but necessary to understanding how change occurs and is institutionalized.

IMPLICATIONS FOR UNDERSERVED POPULATIONS

This is an area that is ripe for study, as little has been reported in the formal literature despite intense interest in the topic. Numerous domestic policy initiatives have been undertaken in the United States, of which the Telecommunications Act of 1996 is perhaps the most well known. UNESCO and the American Association for the Advancement of Science have been active in this area internationally. Moreover, individual publishers (e.g., HighWire Press and ACM) seem to be taking concerns related to access and relative affluence into consideration in developing their pricing and access agreements. But the literature search for this project did not find systematic examination in the literature of either policies, behaviors, or adoption. Admittedly, this research focused only on the literature in English; it is possible that other nations may have investigated these questions but not published the results in this language.

Scientists in major universities in developing nations participate in collaborative activities such as the genetic and protein sequence database initiatives; this behavior is consistent with the international character of scientific research. However, we have not discovered similar inclusive systematic study of the role of e-publication. This issue bears on the question of differential pricing strategies.

INFORMATION SECURITY AND USER PRIVACY

Issues related to information security and user privacy for scientific electronic publishing have not yet been well developed. Behavioral issues relating to passwords are suggested in several studies, but the question of information security does not appear to have been tackled directly. Moreover, although there is a general appreciation of the importance of privacy in the literature, little specific research has been conducted.

At a minimum, a broad understanding of the issues related to information security and scholarly communication, over and above those implied by management of intellectual property rights, is required in order to model the economics of current journal pricing and of archiving. With respect to archiving, for example, the integrity of the archive over time is related to its perceived reliability and hence to the acceptance of electronic journals by some scientific communities.

Understanding the dynamics of electronic journal publishing is challenging because it represents a subset of many larger subjects, from scientific communication to the economics of information goods to information security and networks. The pace of ongoing change within higher education as well as within information technologies will complicate efforts to capture what's happening. These challenges, however, will only serve to make the eventual research that much more valuable.